2012

How discourse context shapes the lexicon: Explaining the distribution of Spanish f- / h- words

Esther L. Brown
Esther.Brown@Colorado.EDU

William D. Raymond
University of Colorado Boulder, william.raymond@colorado.edu

Near-final version (February 2012)

Brown, Esther L. & William D. Raymond (to appear) Diachronica, 29.2, 139-161.

How discourse context shapes the lexicon: Explaining the distribution of Spanish *f-* / *h-* words

Esther L. Brown & William D. Raymond
University of Colorado at Boulder

Using a corpus of Medieval Spanish text, we examine factors affecting the Modern Standard Spanish outcome of the initial /f/ in Latin FV- words. Regression analyses reveal that the frequency of a word's use in extralexical phonetic reducing environments and lexical stress patterns significantly predict the modern distribution of *f-* ([f]) and *h-* (Ø) in the Spanish lexicon of FV- words. Quantification of extralexical phonetic context of use has not previously been incorporated in studies of diachronic phonology. We find no effect of word frequency, lexical phonology, word class, or word transmission history. The results suggest that rather than frequency of use, it is more specifically a word's likelihood of use in contexts favoring reduction that promotes phonological change. The failure to find a significant effect of transmission history highlights the relative importance of language internal sources of change. Results are consistent with usage-based approaches; contextual variation creates differential articulatory pressures among words, yielding variable pronunciations that, when registered in memory, promote diachronic change.

Keywords: phonological change, word frequency, discourse context, language contact, usage-based approach, extralexical phonetic context, Spanish phonology

1. Introduction

Word frequency has long been a linguistic variable implicated in phonological variation and change (Schuchardt 1885, Zipf 1929). Word frequency effects have been reported in a wide range of studies (Labov 1994, Bybee 2001, Phillips 2006, Gries & Divjak forthcoming), with more frequent words typically showing increased rates of reduction compared to lexical items of lower word frequency. Even after controlling such factors as word length, word class, prosodic and lexical stress, phonetic form, speech rate, sociolinguistic factors, and probabilistic measures of language use, many studies have found word frequency to be a predictor of pronunciation variation by at least some measures, especially measures of reduction.

The amount of scrutiny and interest devoted to the effects of word frequency reflects, in part, the theoretical ramifications of evidence corroborating or contradicting such effects. Word frequency effects have been essential components in theories of lexical representation (Bybee 2001, Pierrehumbert 2001) and in theories of language variation (Labov 1981) and language change (Phillips 2006). In fact, so central to contemporary linguistic arguments is word frequency that researchers continue to modify methods of quantification and analysis in order to assure the reliability of reported results, although disagreements continue over what methods are appropriate in assessing frequency effects (File-Muriel 2010, Clark & Trousdale 2009). As Clark & Trousdale (2009) noted, it is imperative not to examine "the effects of lexical frequency on variation

and change in isolation" (38). This admonition follows from the fact that word frequency correlates with any number of other factors, such as word class, word length, register, rate of speech production, and lexical neighborhood density. Given such correlations, the methodological approach chosen might, thus, influence whether or not word frequency effects are found.

In addition to important methodological considerations, the precise nature of word frequency's role in language variation and change has been explored. Bybee (2002) has offered an inspired expansion of this role. In an analysis of word-final *t/d* deletion in American English, Bybee (2002) found that deletion rates of word final /t/ and /d/ in English correspond not simply to word frequency, but to the frequency with which a word with a final /t/ or /d/ is used in specific contexts. Bybee (2002) notes that, "words that more frequently occur in the context favoring a change undergo the change at a faster rate than those that occur less frequently in the appropriate context" (276). For the English word-final *t/d* deletion, an important environment favorable to reduction was found to be pre-consonantal context (Guy 1991: 230, Bybee 2002: 263). The frequency with which each word is used in a specific discourse context that favors reduction (expressed as a proportion) is what we will call a WORD'S FREQUENCY IN A FAVORABLE CONTEXT, or its FFC, and will be a primary focus of this work. There might be a number of contexts that favor reduction, and these contexts will likely affect a word's phonology in different ways. The linguistic factors that can be determined to be favorable or unfavorable to a change will depend upon the aspect of a word's phonology under analysis. Surprisingly, unlike word frequency, the hypothesis proposed by Bybee (2002) regarding the effect of discourse context frequency on variation and change has not been widely tested. Despite broad methodological and theoretical implications, which will be discussed in detail in the following pages, discourse context frequency (expressed as FFC) remains largely unexamined, both in synchronic and diachronic data.

What is the relation between language variation and language change? In usage based theories of language, where phonetic variation in each instance of a word in speech is registered in memory (Bybee 2001, Pierrehumbert 2001, Phillips 2006), context of use can be crucial to studies of phonological variation and change. If a word is used frequently in a discourse context conducive to reduction, it will have more opportunities to reduce, increasing the number of reduced representations registered in the minds of speakers. Conversely, if a word is not used frequently in a context favorable to reduction, it will likely be realized more often as unreduced, so that there will be fewer reduced representations stored. Over time, a new consensus can emerge in the speech community regarding the phonology of initially similar words, with the effects of different contexts of use on different words reflected in a distribution of phonologically distinct lexical forms, conditioned by the reducing environment. In the case of final *t/d* deletion, continued synchronic variation could conceivably eventually lead to changes in lexical representations, with *t/d* being lost in words that occur frequently in contexts with following consonants, but retained in words whose context of use only infrequently places them before a consonant.

The notion that cumulative contextual factors matter is not new. Phonetic context is a nearly universal linguistic factor group considered in analyses of phonological variation. For studies of variation in word-initial position, for example, the potential effects of the phone or phones preceding and following the word-initial segment are used

to measure online articulatory effects. Considering cumulative measures in variation is also not new. Word frequency is an example of a linguistic factor that captures the speaker's cumulative experience with a word. Other cumulative measures include phone and word bigram frequencies (Raymond, Dautricourt & Hume 2006) and phone and word predictabilities (Jurafsky, Bell, Gregory & Raymond 2001). In diachronic studies, cumulative measures are particularly central to the theory of grammaticalization and studies of syntactic change that consider words and constructions in contexts, as well as the frequency of their use in specific contexts (Company Company 2002, Hopper & Traugott 2003, Bybee & Torres Cacoullos 2009).

What is new about our work is twofold. First, despite abundant acceptance that extralexical phonological context significantly constrains variation in production (e.g. Labov 1994, inter alia), to the best of our knowledge, extralexical context is not quantified or included in diachronic studies of phonological change in the same way that probabilistic measures have been done for word-internal contexts (Durie 1996). The theory and methods we propose enable inclusion and quantification of extralexical phonetic context in studies of diachronic phonology through the use of FFC measures. Second, we argue that probabilistic measures of discourse context frequency are, in fact, what the measure of word frequency is capturing through correlation, albeit in a less precise way. Word frequency can act as a measure of discourse context frequency when it is equivalent to FFC, in non-alternating, word-internal environments. In variable, word-external environments, however, a more precise measure of context frequency is required. Using this more precise measure has implications for studies in which no significant word frequency effect could be determined. It is conceivable that an effect of FFC can be found in these cases, making it, thus, of interest theoretically to researchers investigating word frequency. The implications of our work, therefore, are methodological as well as theoretical.

The paper is organized as follows. We first detail previous research that informs FFC and the methods used in the calculation of the measure. We next summarize salient aspects of a problematic development in Romance historical linguistics (F- > [h] > Ø in Spanish) that is the source of the data for the current analysis, in §3. We discuss data and methods in §4, followed by results of our logistic regressions. Lastly, we present a discussion of the theoretical and methodological ramifications of the findings.

2. FFC - Quantifying frequency in a favorable environment

A measure of FFC was initially used in studies of synchronic variation of Spanish /s/. Many Spanish dialects exhibit variability in the pronunciation of /s/, with frequent reduction to [h] or deletion. The use of FFC in these studies was suggested by observations that /s/ reduction variability differs for similar words in ways that cannot be explained by frequency of use. For instance, in New Mexico, where word initial /s/ reduction rates average 18% (Brown 2005a), and in Cali, Columbia, where word initial reduction rates average 17% (Brown & Brown forthcoming), reduction for the lexical item *sí* "yes" is a mere 4% and 1% in the respective varieties. This significantly lower reduction rate for *sí* is found despite the fact that *sí* is far and away the most frequent word with an initial /s/. For example Brown & Brown (forthcoming) note that the word frequency of *sí* is the highest in their data, "with a frequency of 1,774, which compares to frequencies of 975, 782, 397, and 317 for the four next most frequent words: *se, eso, así,*

si" ["3SG and 3PL impersonal clitic pronoun", "demonstrative pronoun *that"*, "adverb of manner/*thus"*, "conditional conjunction/*if"*]. In fact, this lexical item continues to exhibit lower rates of reduction than its lexical frequency would predict, even after statistically controlling for phonological environment and stress. On the other hand, *sí* serves as an excellent example of how discourse context, or FFC, better accounts for variation, as will be illustrated below.

Through an examination of the variable reduction of word-initial /s/ in New Mexican Spanish, Brown (2004, 2006) tests the prediction that discourse context will play a significant role in phonological reductive processes (Timberlake 1978, Bybee 2001, 2002). Aspiration and deletion of word-initial /s/ has been shown through statistical analyses (Brown 2005a, 2005b) to be promoted by preceding non-high vowel (/e, a, o/), presumably owing to the lower tongue height required in the articulation of the non-high vowel. Reduction is inhibited by all other preceding phonological contexts, particularly preceding pause but also preceding consonants and high vowels (/i, u/). Brown (2004, 2006) finds that words with a low FFC (<50% of tokens occur in discourse after a non-high vowel) reduce at a significantly lower rate (6%) than those with a high FFC (25%). Raymond & Brown (forthcoming) test the cumulative effect of such disparate FFC values on words' reduction rates. They confirm that the "… effect of FFC indicates that the cumulative experience of words in reducing phonological contexts … results in a greater likelihood of reduction than context of use alone can explain". That is, even after controlling multiple linguistic factors known to contribute to reduction (i.e.; phonetic context, stress, word frequency, and word predictability), they find effects of the cumulative exposure of words to specific reducing environments (measured by FFC), but not of the cumulative experience with a word overall (measured by word frequency), thus arguing it is not the frequency of use per se that causes reduction, but rather how likely a word is to occur in a reducing environment.

Making this distinction, the low reduction rates of *sí* in Cali, Columbia & New Mexico Spanish can be understood. The discourse context in which *sí* is used in Spanish is overwhelmingly utterance initial position, a phonetic environment not at all conducive to reduction, as shown by studies of /s/ reduction generally (López Chávez 1977, D. Brown 1993, Brown & Torres Cacoullos 2003). Uses of *sí* in utterance initial position are given in examples (1) and (2).[1] The low reduction rate of *sí* in Spanish reflects the propensity for the word to occur in a post-pause environment in discourse, as opposed to the rarer cases in which *sí* fluently follows a non-high vowel (a context conducive to reduction) as in example (3).

(1) *Sí, todos hablan español*.
"Yes, everyone speaks Spanish."

(2) *Sí. Sí es diferente, ¿ve*?
"Yes. Yes it is different, you see?"

(3) *Mi papá sí era muy guapo*.
"My father was (indeed) very handsome."

[1] Data for examples 1-3 form part of the New Mexico and Southern Colorado Linguistic Atlas Project (Bills & Vigil 2008).

It is important to highlight that the low reduction rate of *sí* persists even in phonetic contexts conducive to reduction, such as that illustrated in (3). That is to say, words such as *sí* with a low FFC continue to exhibit significantly lower reduction rates than high FFC words, even in identical favorable phonological contexts, suggesting that the source of a word's reduction from FFC is in its representation and not its articulation.

Synchronic studies indicate, therefore, that FFC may be a better measure of the likelihood of reduction than word frequency, and that an independent FFC effect persists even after bringing under statistical control online articulatory factors such as lexical form and stress (Raymond & Brown forthcoming, Brown 2004, 2006). We assume that mechanisms working to create synchronic variation are the same as those that functioned in the past (Paredes & Sánchez-Prieto Borja 2008), and so presume an FFC effect should be evident in data that reflect historical phonological developments, such as the current distribution in the lexicon of words with phonologically reduced and unreduced variants that are derived from historically unreduced words. Estimates of the discourse context frequency (FFC) of words containing the reduced and unreduced variants can be made using corpora of historical documents by identifying independent linguistic factors known to inhibit or promote change (such as specific extralexical phonological contexts). The following describes the historical sound change and the resultant contemporary lexical distribution of forms that we investigate in the present study to test the application of the FFC measure to diachronic data.

3. Spanish F- > [h] > Ø

A much-examined issue in Spanish historical phonology is the outcome of Peninsular Latin words with initial /f/ (F- words) in Modern Standard Spanish (MSS) (see Menéndez Pidal 1926-1968: 198–208, Baldinger 1972:18–22). When Peninsular Latin F- was followed by a liquid or a glide (FC- words, e.g., *flor* "flower", *fuente* "fountain"), these words predictably appear with [f][2] in MSS. However, when followed by a vowel (FV- words), the reflex of the initial [f] is indeterminate solely from the word's phonology. As in (4) and (5), the sound may be either lost completely (*hablar* < Lat. *fabulari* "to talk") or may also be [f] in MSS (*favor* < Lat. *favor* "favor").

(4) *F-* > … > [Ø] / __ [vowel]
(*hijo* "son", *hecho* "done", *hallar* "to find", *horno* "oven", *humo* "smoke")

(5) *F-* > … > [f] / __ [vowel]
(*fijo* "fixed", *fecha* "date", *favor* "favor", *foco* "focus", *fumar* "to smoke")

Much of the literature on the distribution of the reflexes of the initial [f] of F- words in Spanish has been dedicated to attempting to account for precisely the outcomes described in (4) and (5). That is, why was the initial consonant in some FV- words lost, while in others it appears as [f]? Given the examples in (4) and (5), plainly MSS outcomes are not determined by the following vowel, and, MSS phonology also cannot be determined from

[2] There is much social and dialectal variation regarding allophones of /f/, with a common variant of /f/ being the voiceless, bilabial fricative [ϕ] (Lipski 1995: 285–286). The results we report can be extended to these varieties as well.

any specific Peninsular Latin etymon, given examples such as *humo* "smoke" and *fumar* "to smoke", which are both derived from the same Latin root *fum-* "smoke".

Attempts to account for the phonology of the words in (4) and (5) typically focus on the histories of these words. Broadly, it is widely accepted that phonological development of Ibero-Romance resulted in the substitution of the fricative /f/ in all words by /ɸ/ (Penny 1991), perhaps because of substrate influences of languages on the peninsula that lacked /f/ (Menéndez Pidal 1964: 198–233). The fricative /f/ was reintroduced into medieval Spanish in the thirteenth century by speakers of languages such as French and Gascon as well as through peninsular dialects that had maintained /f/. The phoneme /f/ subsequently supplanted /ɸ/ in all environments. The appearance in MSS of [f-] in some FV- words but not others has then been accounted for by differential histories of these words. Relevant historical paths include the introduction during the Medieval period by /f/ speakers of FV- words that had been absent from the transmitted vocabulary (referred to in Spanish as *cultismos*; e.g., *favor* "favor", *feroz* "fierce", *figura* "figure", *forma* "shape, form", *futuro* "future"), and the reintroduction of [f-] pronunciations to some FV- words in the transmitted vocabulary (which had previously undergone reduction) through the influence of /f/ speakers (referred to as *semi-cultismos*; e.g., *falso* "false", *fe* "faith", *fin* "end, limit", *fondo* "bottom (of something)", *fundir* "to found (metal)"). Evidence for the re-introduction of [f-] comes from forms such as *febrero* "February" and *feria* "celebration", which also have 12th century popular or rustic attestations (*hebrero* and *heria*, respectively) (Blake 1988: 53). Language-internal sources of phonological adjustment, including analogy, homophony avoidance, and hypercorrection, have also been adduced to explain [f-] in some orally transmitted words in which MSS [f-] is otherwise unexpected (e.g., Penny 1991). Thus, FV- words with a continuous oral tradition from Latin will largely appear with no onset in MSS, although some orally transmitted words are assumed to have had [f-] restored. FV- words without a documented history preceding Medieval Spanish are all assumed to have been introduced in the medieval period from other speech sources with [f-]. Note, thus, that [f-] in MSS is taken as evidence of their non-oral source. Such explanations have been deemed circular (Blake 1988: 53) and call into question the explanatory power of the term *cultismo* (Badía Margarit 1972).

For most FV- words with [f-] in MSS explanations based on source histories are not controversial because there is a clear record of their introduction into Medieval Spanish (e.g., *familia* "family", first attested in 1220-1250). For some FV- words with [f-] in MSS, explanations become more idiosyncratic and less convincing because their early attestation suggests that they were orally transmitted from Latin and not introduced. Either these words never lost [f-] or [f-] was restored. In either case, the question is why these words (and others like them) have [f-]. Two words cited by Penny (1990: 178) perhaps illustrate this point. Both *feo* "ugly", first attested 1140, and *fino* "fine", first attested in the early thirteenth century, do not appear to be either *cultismos* or influenced by prestige or dialect pronunciations (*semi-cultismos*). Penny (1990: 178) suggests possible explanations for the persistence of [f-] in these words as homophony avoidance for *feo* (avoiding FETA > *heda*, "woman who has recently given birth") and analogy with *fin* for *fino*. Although such mechanisms exist, their application to isolated words historically may be unfalsifiable and, at best, only narrowly applicable.

Given that the diachronic trajectory of FC- words can be accounted for by lexical phonology (i.e., the sound structure of the words themselves, specifically the class of the phone following [f-]), we suspect that the difference between (4) and (5) might also depend more generally upon phonological factors to some extent. In addition to the FC- words, phonological context has been shown to be important in the development of [f] in Spanish and other Romance varieties (e.g., Gascon, some Italian dialects, and Sardinian) where [f] is retained only post-consonantally in word-medial position (Naro 1972, Pensado 1993).

Phonetic conditioning has clearly played a role in the development of Latin F- in Spanish, as plainly demonstrated by the FC- words. When the initial consonant precedes [l, r, w, j], it remains [f]. Nearly all treatments of the development of Latin F- note that the diachronic change has been conditioned by the segment or phone following the word-initial [f] (Menéndez Pidal 1926-1968: 221, Naro 1972: 443, Blake 1987: 71, Penny 2000: 71). For example, Penny (1991) outlines the loss of F- in Spanish as the natural outcome of normal allophonic variation "governed by the nature of the following phoneme" (80) and Blake (1987: 71) states that "la F- Latina se realiza o bien como fricativa aspirada [h] o bien como elemento nulo [ø] siempre que va seguido de un segmento silábico" ("Latin F- is realized as the aspirated fricative [h] or the phonologically null [ø] whenever it is followed by a syllabic element"). Clearly lexical phonology has been a factor. Although the role of a vowel following an initial [f] is not determinative, aspects of vowel quality may similarly influence the outcome.

Consequently, the effect of the phonological context following the [f] has been recognized, but the effect of the extralexical preceding phonological context has been examined to a much lesser degree as a conditioning factor for the distribution in word-initial environments. As we have suggested, this methodological lapse is true not just for Latin F- in Spanish, but for studies of diachronic phonology generally. For medial [f], Pensado (1993) notes that in the adaptation of Arabic loanwords in Spanish, "Por lo que respecta a la *f*, hay un resultado *h* en posición intervocálica. En posición postconsonántica y cuando es geminada, la *f* se conserva" (154) ("With regards to the *f*, one outcome is *h* in intervocalic position. In post-consonantal position and when a geminate, the *f* is maintained").

Preceding context is thus a significant factor in the phonological development of some word-medial [f] tokens. The preceding context of a word's initial phone is not constant, but distributional statistics can be collected for each word that quantify the tendency for the word to appear in the company of certain other words, and hence sounds. Penny (1972: 466) and Menéndez Pidal (1926-1968: 213) note such distributional tendencies might play a role in the phonological development of a few lexical items in Spanish (e.g., the toponyms *Porma*, tending to occur after *en* (IN FORMA), and *San Fagún* > *Safagún* > *Sahagún*). There is also some evidence for the importance of contextual distribution from other languages. Naro (1972) mentions that in Bitti, a province of Sardinia, there is a shift from [f] to [h] if the word-initial [f] is in "intervocalic position in the speech chain" (444). Hall (1968), looking at syntactic doubling in Western Romance, posits that position in the breath group likely plays a significant role in word-initial reduction.

Though improbable to some (e.g., Pensado 1993), many have drawn parallels between the historical development of F- and the synchronic variation demonstrated with

syllable-initial [s] in Spanish (Malmberg 1958, Mason 1994, Méndez Dosuna 1996). Both voiceless fricatives can undergo phonological reduction in syllable-initial position in some dialects, and it has been noted that "there is little doubt that /s/ > [h] and /f/ > [h] are instances of weakening with a clear articulatory motivation" (Méndez Dosuna 1996: 99). An articulatory explanation of the lenition of /f/ to [h] is provided in Foulkes (1997:264), who summarizes:

> If speakers were to assimilate the articulation of the labial fricative to adjacent vowel sounds, that is sounds in which the mouth aperture can be relatively large, then the labial opening during the fricative articulation might be increased to such an extent that labial frication is either significantly reduced or wholly impossible. By removing oral frication in this manner, the result would be a voiceless vowel, interpretable as [h].

Adjacency to non-high vowels could then promote f > [h], as it has been shown to do for /s/ > [h].[3]

Given the parallel articulatory motivation and pathways of change for both F- and [s], we hypothesize that many of the same factors shown to be relevant for [s] in previous analyses (Brown & Torres Cacoullos 2003, Brown 2004, 2006, Raymond & Brown forthcoming), such as lexical stress, following phonological context, word frequency, and discourse context frequency (FFC), would also be important in the development of F-. Indeed, if we consider preceding non-high vowels to be a discourse context conducive to reduction, it appears as though 13th century usage patterns differed for MSS *h*- and *f*- words. A calculation of usage patterns in Davies (2002-) 13th century corpora of all *h*- words listed in Corominas (2006) derived from Latin FV- words (30 words in Corominas, 27,213 tokens in Davies) and all *f*- words derived from orally transmitted Latin FV- words listed in Corominas (2006) (9 words in Corominas, 1,862 tokens in Davies), reveals the *h*- words have, on average, a higher FFC (.53) than *f*- words (.35). That is to say, based upon a small set of words, without taking other factors into account, *h*- words seem to have been used in discourse contexts that were more favorable to reduction (i.e.; used more frequently in a post non-high vowel discourse context).

A comparison of *f*- and *h*- word pairs that share an etymon, and hence share following vowel and stress patterns, is similarly suggestive of a potential FFC effect. FFC calculations of 12th–14th century texts (Davies 2002-) reveal that, on average, when *h*- words are compared to their *f*- doublet, the *h*- words have a higher FFC (.44) than the *f*- counterpart (.38), albeit *h*- words do not uniformly have higher FFC values. This information is summarized in Table 1.

[3] Foulkes (1997: 264) also proposes an acoustic explanation and argues that although the articulatory explanation "is in itself plausible" it is not sustained in his cross-linguistic survey, "where /f/ > /h/ is found in a restricted set of environments, it is always in the context of a high back rounded vowel". Foulkes (1997) does not consider preceding phonological context. Results of our analyses do not reveal following /u/ as favorable in the FV- change.

Table 1. Calculated FFC average of *f-* and *h-* doublets derived from Latin FV- words in 12-14th century texts (Davies 2002-).

Latin	MSS *h-* word	FFC	MSS *f-* word	FFC
fundus	hondo "deep"	.28	fondo "bottom"	.22
forma	horma "shoemaker's last"	.44	forma "form"	.59
facies	haz "visage"	.61	faz "face"	.47
filum	hilo "thread"	.18	filo "edge"	.18
factus	hecha "done"	.53	fecha "date"	.39
fumare	humear "to emit smoke"	.40	fumar "to smoke"	.24
fundere	hundir "to submerge"	.62	fundir "to found (metal)"	.60
Average		**.44**		**.38**

Comparisons of these *h-* and *f-* data sets, which do not control for any probabilistic relationships or other factors that may influence the MSS form of FV- words, suggest the presence of an FFC effect. The following section outlines the model we design to test for the potential effects of FFC while controlling for other linguistic factors such as word frequency and stress.

4. Data & methods

To examine the effects of statistical distributional patterns and phonological variables on the modern distribution of [f] vs. [Ø] in FV- words, we used as the dependent variable arguably the least controversial piece of the F- > [h] > Ø historical development; the modern Standard Spanish (MSS) outcome. That is, in MSS, is the FV- etymon realized with a word initial [f] (spelled *f)* or Ø (spelled *h*)?

Our data were taken from the historical text *La Celestina* (Fernando de Rojas), written during a time of much phonological (and orthographic) instability in FV- words (Penny 1990). The historical text was not used to determine the dependent variable; that is, the text was not used in any way to determine precise realizations of FV- words at the time it was written or the precise orthographic significance of a written *f-*, *ff-*, *h-*, or Ø. Use of the historical text was important in accounting for word order differences between 15th and 21st century Spanish. For example, Penny (1991:123) notes unstressed pronouns in Old Spanish "followed rules different from those of the modern language" with regard to positioning, which we illustrate in example (6).

(6) *maldito seas, que <u>fecho me</u> has reyr* [*La Celestina*]
maldito seas, que <u>me</u> has <u>hecho</u> reír (MSS)
"damn you, you made me laugh"

Such word order differences between the centuries could influence word context distributions. The text was also used to estimate probabilistic word and phone measures, because word frequency values might also be expected to vary across centuries (Sonderegger & Niyogiy 2010). For instance, the word frequency of the verb *hallar* "to find" was 413 per million in the 13th century, but had fallen to 91 per million by the twentieth century, while the frequency for the verb *faltar* "to lack" rose from 7 per million in the1200s to 118 per million in the last century (Davies 2002-).

The *La Celstina* text contains 66,000 words and provided us with 1,848 word tokens derived from FV- (written in the text with either *f-* or *h-*). We excluded from the data all FC- words, which have a uniform diachronic development to [f] in MSS, as noted already. Each word type was considered a separate item, as opposed to conflating all paradigmatically related forms into one lemma (e.g., all forms of *hallar* "to find", such as 3SG IND. *halla*, 3PL IND. *hallan*, 1SG or 3SG IMP. SUB. *hallara*, 1SG or 3SG COND. *hallaría*, are considered separate types.) Of the resulting 346 types, 41 had tokens that varied orthographically within the text (i.e., *f-/h-* variation, e.g., *fablar*/*hablar* "to talk"). The orthographic variation does not play a role in our analysis, because it is the MSS pronunciation that we are predicting. Tokens of orthographically variable words (e.g., *fablar* and *hablar*) were counted as one type, despite the different spellings. Analysis was performed on the 346 types.

The tokens were coded for preceding and following phonological context. The contexts immediately preceding the word-initial consonant were classified into two categories, based on the orthography of the text: non-high vowels (/a, e, o/) and all other contexts (i.e., high vowels & glides (/i, u, w, j/), consonants (/n, l, r, s, d/), and pause in cases of sentence-initial *f-* or *h-*). The following phonological context was coded and categorized into non-high vowel (/a, e, o/) and high vowel (/i, u/).[4] The lexical stress of each word was also coded, indicating whether the word-initial consonant formed part of a lexically stressed or unstressed syllable.

Word frequencies for FV- words were calculated from counts of tokens in the *La Celestina* text, and values were normalized to frequency per million. These frequencies are taken to be representative of usage patterns of the time. The historical text was also used to calculate FFC for each type by taking the proportion of tokens in the reducing environment out of the total number of tokens of each type. Based on findings for the New Mexican syllable-initial [s] reduction data, contexts considered favorable to reduction were those containing a preceding non-high vowel (/a, e, o/). These contexts were contrasted with all other preceding phonological environments (i.e., high vowel, diphthong, consonant, and pause). FFC is defined for each word type as the proportion of instances of a word following a non-high vowel out of the total number of tokens of that word that appeared in the text.

The FFC values of words differ considerably. Some words frequently occur in contexts conducive to reduction, such as the word *fazer* (MSS *hacer*, "to make, to do"), seen in example (7). This word occurs 74 times in our corpus. In 55 of those cases, the word-initial consonant is preceded by a non-high vowel, as is illustrated in 7a-c. In the remaining 19 instances of use in *La Celestina*, *fazer* follows a segment other than /a, e, o/, as is illustrated in 7d. In 74% of the cases, therefore, this word is used in a phonological context conducive to reduction, giving it an FFC value of .74.

(7) a. *... me lo as vist**o f**azer ...* [Celestina]
"… you have seen me do it …"
b. *… has pensado d**e f**azer esta piedad?* [Celestina]
"… you have thought to do this charity?"

[4] Penny (1991) suggests that it was rounding of the following vowel that was determinative in the change. We test rounded [o, u] vs. unrounded [i, e, a], as opposed to high [i, u] vs. non-high vowel [e, a, o], but found no effect.

c. … *que se podr**a f**azer sobre* … [Celestina]
"… what can one do about …"

d. … *lo que he visto a mucha**s f**azer* … [Celestina]
"… what I have seen many do …"

Other words, conversely, are often found in contexts that do not favor reduction, such as *fin* "end" in example (8), which occurs 51 times in our corpus, 17 in the non-high vowel context illustrated in 8a. The remaining 34 cases of *fin*, on the other hand, occur outside of the favorable /e, a, o/ extralexical context, as illustrated in 8b-e, giving *fin* an FFC value of .33.

(8) a. … *dand**o f**in a tus día s*… [Celestina]
"… putting an end to your days…"
b. … *e**l f**in de tu deseo.* [Celestina]
"… the end of your desire."
c. *sin esperanza de bue**n f**in.* [Celestina]
"… without hope of a good end."
d. … *mi fin es llegado* … [Celestina]
"… my end has arrived."
e. *quieres pone**r f**in a tu vida* [Celestina]
"you want to end your life"

Not only does the FFC of individual words vary, but the contextual (textual) use of the *f*- and *h*- word classes varies significantly. FV- words in *La Celestina* with a low FFC are about equally split between *h*- words and *f*- words, as can be seen illustrated in Figure 1. However, words with a high FFC are predominantly *h*- words as opposed to *f*- words. Discourse context frequency (FFC) is included in the analysis as a separate, independent linguistic variable.

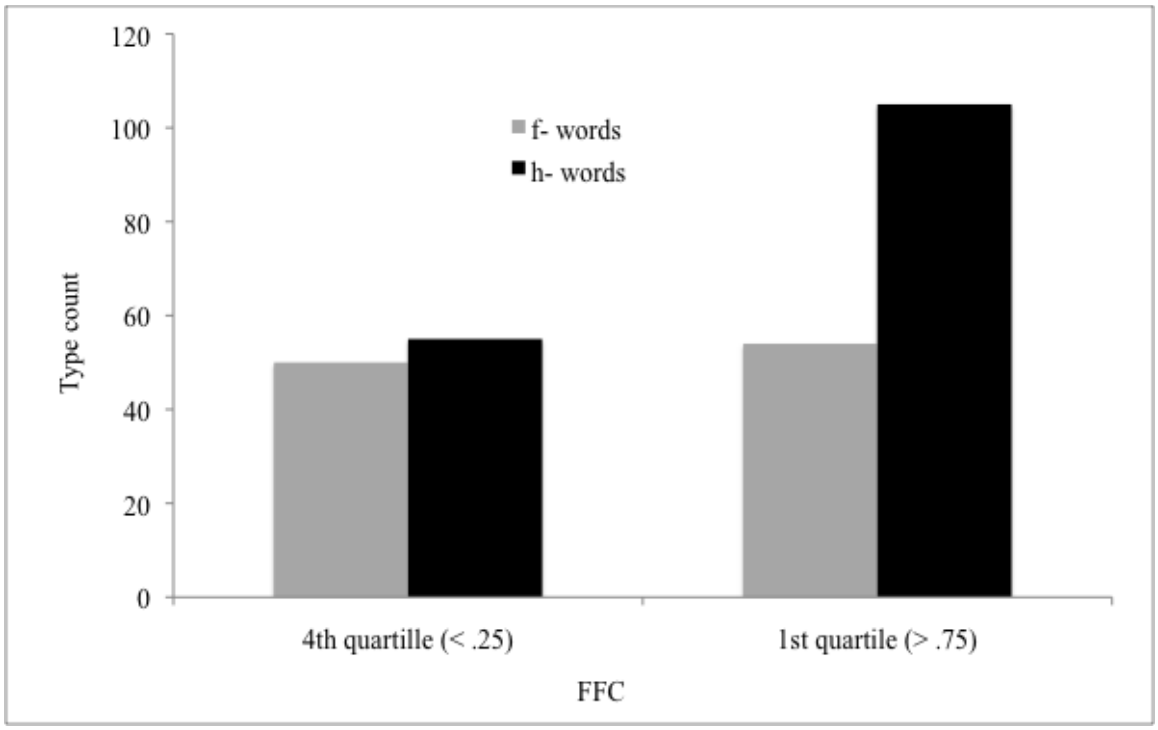

Figure 1. Number of *f*- and *h*-word types in the *La Celestina* corpus with low and high FFC.

Because FFC is a cumulative measure, more representative values are likely obtained from types with more than one token in the corpus. To account for the potentially skewed FFC value obtained from a single example of a word (e.g., FFC of zero, e.g., *ferocidad* "ferocity" [muy gozosa con su ferocidad], *fantasía* "fantasy" [trae a mi fantasia la presencia angelica], *hervía* 3SG IMP. "to boil" [la sangre que ayer heruia], *herreros* "iron workers" [si va entre los herreros]; or FFC of 1.0, e.g., *formada* "formed/shaped" [como sea de carne sentible formada], *forzaste* 2SG PRET. "to force" [tu le forzaste a darle fe], *hormiga* "ant" [como hormiga que dexa de yr], *hoyo* "hole"[o si ha caydo en alguna calzada o hoyo]), we coded hapax tokens: The hapax variable is *yes* if a token occurred only once in the text and *no* if it occurred more than once.

Finally, each type was also coded for method of word transmission (oral, non-oral). In an attempt to avoid circularity in determining transmission type (i.e., calling a word with an [f-] non-oral), any word with continued use from Latin to Spanish was labeled oral. Note that orally transmitted words thus include those classified by Penny (1990) as *semi-cultismos*. Words listed in Menéndez Pidal (1926-1968), Penny (1990, 1991), and/or Corominas (1961-2006) as *cultismos* were labeled non-oral.

Table 2. Variables coded for each token in the FV- word dataset.

1. Realization of initial phone in MSS (*f*- = [f-] = *unreduced*; *h*- = Ø- = *reduced*)
2. Proportion of times in the corpus that the FV- word has a preceding context favorable for fricative reduction, which is the proportion of tokens for an FV- word type that are preceded by a non-high vowel (Frequency in a Favorable Context, or FFC);

3. Favorability of the phone following F- for reduction. We coded *yes* for following non-high vowels (e.g., *fabricar* "to create, to make", *felicidad* "happiness", *fortaleza* "strength", *hablar* "to talk/to speak", *hecho* "past participle to make/to do", *hormiga* "ant") and *no* for all other following phones (e.g., *fin* "end", *furia* "fury", *hígado* "liver", *humo* "smoke");
4. Log of FV- word frequency per million;
5. Hapax words in the corpus (*yes* for words with only a single token in the dataset, such as *fácil* "easy"; *no* for all words with more than one token, such as *harán* "3PL.FUT. to make/to do");
6. Stress on FV- syllable (*stressed* if lexical stress on primary syllable (*hablas* "2SG speak/talk"), *unstressed* if lexical stress on non-initial syllable (*hablaremos* "1PL speak/talk");
7. Mode of transmission (*non-oral* if listed in etymological dictionaries as learned, *oral* if transmitted from Latin).

5. Results

We performed analyses on the type dataset with the R statistical package using logistic regression, with MSS outcome as the dependent variable; reduced initial Ø (e.g., *humo* "smoke") vs. unreduced initial [f] (e.g., *fumar* "to smoke"). A logistic regression models factors that can be shown to predict a distribution, but only probabilistically. Using such a model, a phonological form cannot be determined from a number of contextual factors in a rule-like way. We included measures 2-7 listed in Table 2 and their pairwise interactions. The model likelihood ratio for the model was 134.06 (d.f. = 10; $p < .001$). Classification accuracy was 83.8% compared with a simple model in which all tokens are assumed to be reduced, which would only accurately classify 74% of the tokens. The overall correlation for the model was Nagelkerke $r^2 = .44$. The results of the analysis of the complete dataset are shown in Table 3, along with the odds ratios for the significant predictors.

Table 3. Results of analysis of the complete dataset (N = 346).

Variable (application value)	*p*	Odds Ratio Effect
FFC (higher)	0.027	1.55
stress (unstressed)	0.014	1.55
hapax (yes)	0.055	1.35
stress x hapax	0.067	1.26

As shown in Table 3, there was a main effect of FFC. The discourse context frequencies of FV- words (FFC) significantly predicts the MSS lexical distribution of initial [f] and Ø. Words used in discourse more frequently in a reducing environment (in this case, a post non-high vowel context) are 1.55 times more likely to be realized as Ø in MSS (e.g., *hablar* "to talk/speak", *hacienda* "hacienda", etc.) than words more commonly used in non-reducing environments (e.g., *fino* "fine", *falso* "false, fake", etc.). In fact, overall the *f*-words had an average FFC of .44 and the *h*- words had an average FFC of .61. There was also a main effect of the stress variable on reduction in the dataset, with no stress on the initial syllable of an FV- word making it 1.55 times more likely to be reduced in MSS than if the initial syllable is stressed. This result is in line with the

general pattern noted by Guy (2005), that "syllabic stress gives greater prominence to the features of a syllable, and favours their retention" (375). There was a marginal effect (p .055) in the data for hapax words. Words that occur just once in the data are more likely to be Ø in MSS (*h*- words) than tokens with multiple examples.

Although we ran all pairwise interactions in the dataset, only one marginally significant interaction was found in the analysis; stress and hapax. The interaction reflects the fact that hapax words behave differently with regards to stress. For words that occur more than once, the reduction pattern is what would be expected from the main effect of stress (more reduction in unstressed syllables). However, in words that occur only once, it is the opposite. The stressed hapax words, which account for only 10% of the data, are more reduced. Note that there were no significant interactions involving FFC.

The significant result for FFC was obtained, recall, even when considering other factors with potential explanatory power. Importantly, transmission history was not found to be significant, nor was word frequency or following vowel class. It is noteworthy that not only were these variables not able to predict MSS, but there were no significant interactions in the data between any of these variables, such as between word frequency and transmission history (i.e.; low word frequency and non-oral words) or word frequency and hapax (i.e.; low word frequency and hapax).

6. Discussion

The failure to find an effect of transmission history in our data was surprising. This finding was perhaps particularly striking given the general bias in the literature on the topic of F > h > Ø in Spanish. As noted by Blake (1988), assuming regularity of sound change, in the face of the lexical irregularity in the MSS outcome (e.g., ***f****umar* "to smoke" but ***h****umo* "smoke"), language external explanations are often adduced. Further, an adequate and exhaustive list of phonological conditioning environments for the F- > [h] > Ø change, hitherto unidentified, placed considerable emphasis on finding a language external (contact) explanation. The results of our logistic regression highlight language internal sources for the MSS outcome of *f*- and *h*- words (FFC, stress). Clearly, we do not deny any role for dialect and language mixture with these data, particularly in light of such sociocultural evidence that exists on the subject (Menéndez Pidal 1968, Penny 2000, Spaulding 1943). Nor do we advocate a monocausal explanation of this change. While acknowledging the importance of the effects so often noted in the literature from the explanation of MSS *f* and *h* (analogy, homophony avoidance, dialect contact, language contact), our result argues for the primacy of language internal linguistic factors EVEN IN LANGUAGE CONTACT SITUATIONS, a point also highlighted by others (Sanchez 2008). In this case we find a significant and relatively weighty impact of FFC in this particular phonological outcome. A particularly problematic aspect of traditional accounts of F- > h > Ø has been the circularity of labeling all *f*- words in Spanish derived from Latin FV- as learned, while identifying them as learned based on the presence of the *f*-. Our result suggests that before attributing an external explanation, that language internal patterns be thoroughly explored.

What is the relationship between FFC and *f*-/*h*- outcomes in MSS? Based on the FFC results presented in Table 3, we could argue that regardless of the transmission history of the FV- words, the discourse patterns would similarly influence the pronunciation. Words participating in the reductive process through the period of [f-] loss

(orally transmitted FV- words), must have changed at different rates significantly influenced by the discourse contexts in which they were used. The result was that perhaps not all FV- words completely lost their initial [f-]. This conclusion is very much in line with Durie (1996:131), who notes

> that lexically gradual yet phonologically conditioned sound change, as it has been observed in progress by Labov and others, will not necessarily be exceptionless in its final outcome. Such instances can be regarded as true cases of 'sound change', even if their final outcome is not categorically regular in the classical sense.

This state of lexical variation existed at the time of language contact during which the new *f-* initial words were introduced by prestige speakers. Reintroductions of words beginning with [f-] would have the effect of strengthening the variable words already in the language. These reintroduced words would themselves be subject to discourse context pressures, which might have operated selectively on which words were borrowed, as suggested by the lack of significant interaction in our data between FFC and transmission history. In our data, non-oral *f-* words have an FFC almost identical to that of oral *f-* words (.51 and .50 respectively), which is to say that irrespective of transmission history, the two classes of words were used in discourse contexts that did not differ significantly [$t(126) = .40$, $p = 0.69$)]. Conceivably native speakers were more amenable to borrowing *f-* words that were used in low FFC contexts.

However, apart from finding an overarching pattern for this issue in Romance phonology, this study establishes the cumulative of extralexical conditioning contexts (as measured by FFC) as a significant and independent force in phonological variation and change. We know this now for synchronic variation (Bybee 2002, Brown 2004, 2006, Raymond & Brown forthcoming) and, as illustrated in Table 3, for diachronic phonological development, as well. This was perhaps an anticipated result, since, as Guy (2005) notes, "the processes and mechanisms of diachrony should be reflected in synchronic variation" (372). Models of lexical representation and theories of language change, therefore, must be able to account in some way for the variation that registers these specific usage patterns. As Bybee (2002) notes, one model able to capture the phonologically and lexically gradual nature of this variation and change is the Exemplar Model of lexical representation (Bybee 2001, Pierrehumbert 2001).

What are the implications of such a result? As noted by Labov (1994:460), we know these usage-based patterns of variation and change are often attributed to word frequency. Results and studies summarized in this work suggest that rather than word frequency, it is FFC, or more generally the likelihood that a word occurs in reducing environments, that we should calculate. This conclusion may help account for the lack of consistent word frequency effects cited in the literature. For example, Barras et al. (2007:8) note that predictions,

> made by proponents of usage based theories such as Bybee, do clearly hold for some data, but as we show … they are not universal patterns. We therefore concur with Labov (2006), who suggests that frequency effects are not universal: while phonological changes may display lexical and social effects, often the main constraint to change is phonetic environment.

Our results argue that it is the *cumulative* effect of phonetic environment and not exclusively the *online* effect of phonetic environment that constrains variation. Further, methodologically, FFC as a factor allows for the calculation of extralexical effects on phonological developments. In this way, FFC can be particularly relevant to diachronic analyses of sound change, especially for changes at word boundaries.

The explanation for how the FFC of preceding context provides a better measure of reduction than word frequency in our study may lie in the measure itself. Jurafsky, Bell, Gregory & Raymond (2001:233), report that factors that combine "several independent measures", such as conditional probability which merges "joint probability and the relative frequency of the neighboring word", provide a more accurate picture of the reductive process. In this way, FFC, which measures word frequency *in combination with* independent conditioning factors (such as extralexical preceding/following phonological environments), is a more precise measure and predictor of change than word frequency alone.

However, beyond a methodological refinement, these results additionally help to specify what word frequency effects are. It does not seem to be the case, as we illustrated previously with *sí* in Spanish, that reductive change has a direct and unavoidable correlation to word frequency. As Raymond & Brown (forthcoming) note, "an explanation based simply on how often a word is used would seem to entail that reductive change should occur uniformly across the word and not merely on certain segments or syllables, contrary to observations of lexical change". We conclude that word frequency effects are not context independent, but rather register effects of specific reducing environments on each lexical item.

The effect of discourse context may have interesting correlations with the word class effects noted by Phillips, and to a primary question she addresses, "Why do some word classes undergo a sound change before others?" Phillips (2006:96) notes that function words typically undergo weakening processes first, "probably due to their low sentence stress" (2006:112). This frequency of use in non-prosodically prominent position we would argue could be an FFC effect, and in fact, word class effects may be manifestations of differential patterns of use. If the various discourse contexts differ in the degree to which they favor or disfavor a specific change, we would predict word class effects of the type summarized in Phillips (2006). It is the cumulative effect of specific reducing environments that may account for word class effects. The dissimilar patterns of use between different grammatical categories could also shed new light on the study of homonymic pairs. It is likely that homonyms, particularly if belonging to diverse word classes, are used in appreciably different discourse environments, which as we have shown, has a significant effect on reduction rates.

The type of FFC we have studied measures the cumulative articulatory effect of a reducing environment in the context immediately preceding the word, which varies across uses. However, words also vary in their probability of occurrence in environments that encourage reduction through other mechanisms. As previously mentioned, factors known to promote reduction include faster speech, lack of prosodic salience, and greater predictability. It remains to be determined whether FFC based upon measures of these factors might also be found to correlate significantly with reduction.

We have shown that one measure of the cumulative effect of a word's occurrence in a reducing environment, which we have called FREQUENCY IN A FAVORABLE CONTEXT, is an independent, and at times powerful predictor of reductive variation and change. It is certainly not the only factor involved in reduction, and the magnitude of effect will likely vary depending upon the phonological reductive process examined. More studies on different processes and different languages are needed to confirm the efficacy of FFC in explaining variation and change.

References

Badía Margarit, Antonio. 1972. "Por una Revisión del Concepto de "Cultismo" en Fonética Histórica". *Studia Hispánica in honorem R. Lapesa I. Madrid* ed. by Menéndez Pidal, 137–152. Madrid: Cátedra.

Baldinger, Kurt. 1972. *La Formación de los Dominios Lingüísticos en la Península Ibérica*. Madrid: Gredos.

Barras, Will, Patrick Honeybone & Graeme Trousdale. 2007. "Is word-frequency a factor in phonological demergers?". Paper presented at Second International Conference on the Linguistics of Contemporary English 2-4 July, 2007, University of Toulouse-le Mirail. http://www.ling.ed.ac.uk/~s9527813/docs/bht-toulouse.pdf

Bills, Garland & Neddy Vigil. 2008. *The Spanish Language of New Mexico and Southern Colorado: A linguistic atlas*. Albuquerque, NM: University of New Mexico Press.

Blake, Robert. 1987. "Aproximaciones Nuevas al Fenómeno de [f] > [h]> [ø]". *Actas del I Congreso Internacional de Historia de la Lengua Española* vol. I ed. by Manuel Ariza, Antonio Salvador, Antonio Viudas, 71–82. Madrid: Arcos Libros.

Blake, Robert. 1988. "Sound change and linguistic residue: The case of [f-] > [h-] > [ø]". *Georgetown University Round Table on Languages and Linguistics* ed. by Thomas Walsh, 53–62. Georgetown, DC: Georgetown University Press.

Brown, Dolores. 1993. "El Polimorfismo de la /s/ Explosiva en el Noroeste de México. *Nueva Revista de Filología Hispánica.* 1.159–176.

Brown, Esther L. 2004. *Reduction of Syllable-initial /s/ in the Spanish of New Mexico and Southern Colorado: A usage-based approach*. PhD Dissertation, University of New Mexico.

Brown, Esther L. 2005a. "Syllable-initial /s/ in traditional New Mexican Spanish: Linguistic factors favoring reduction 'ahina'". *Southwest Journal of Linguistics* 24.13–30.

Brown, Esther L. 2005b. "New Mexican Spanish: Insight into the variable reduction of 'la ehe inihial' (/s-/)". *Hispania* 88:4.813-824.

Brown, Esther L. 2006. "The effects of discourse context on phonological representation in Memory". Paper presented at the 80th Annual Meeting of the Linguistic Society of America (LSA). Albuquerque, NM, January 5-8.

Brown, Esther L. & Rena Torres Cacoullos. 2003. "Spanish /s/: A different story from beginning (initial) to end (final)". *A Romance Perspective in Language Knowledge and Use. Selected papers from the 31st linguistic symposium of Romance languages (LSRL)* ed. by Rafael Núñez-Cedeño, Luis López & Richard Cameron, 22-38. Amsterdam: Benjamins.

Brown, Earl & Esther L. Brown. Forthcoming. “Syllable-final and syllable-initial /s/ reduction in Cali, Colombia: One variable or two?”. *Linguistic Studies in Colombian Varieties of Spanish* ed. by Richard File-Muriel & Rafael Orozco, Madrid: Iberoamericana.

Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge, UK: Cambridge University Press.

Bybee, Joan. 2002. “Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change”. *Language Variation and Change* 14.261–290.

Bybee, Joan & Rena Torres Cacoullos. 2009. “The role of prefabs in grammaticization: How the particular and the general interact in language change”. *Formulaic Language Volume 1. Distribution and Historical Change* ed. by Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali & Kathleen M. Wheatley, 187–217. Amsterdam: Benjamins.

Clark, Lynn & Graeme Trousdale. 2009. “Exploring the role of token frequency in phonological change: Evidence from TH-fronting in East-central Scotland”. *English Language and Linguistics* 13:1.33–55.

Company Company, Concepción. 2002. “Grammaticalization and category weakness”. *New Reflections on Grammaticalization* ed. by Gabriele Diewald & Ilse Wischer, 201–215. Amsterdam: Benjamins.

Corominas, Joan. 2006. *Breve Diccionario Etimológico de la Lengua Castellana*. 3a Edición Madrid: Gredos.

Davies, Mark. 2002-. Corpus del Español (100 million words, 1200s-1900s). Available online at http://www.corpusdelespanol.org.

Durie, Mark. 1996. “Early Germanic umlaut and variable rules”. *The Comparative Method Reviewed: Regularity and irregularity in language change* ed. by Mark Durie & Malcolm Ross, 112–134. Oxford, UK: Oxford University Press.

Fernando de Rojas. 1993. *La Celestina* ed. by Dorothy S. Severin. Madrid: Cátedra, 1993. Project Gutenberg Etext. 1998 R.S. Rudder. http://www.gutenberg.org/cache/epub/1619/pg1619.txt

File-Muriel, Richard. 2010. “Lexical frequency as a scalar variable in explaining variation”. *The Canadian Journal of Linguistics/La revue canadienne de linguistique* 55:1.1–25.

Foulkes, Paul. 1997. “Historical laboratory phonology – Investigating /p/ > /f/ > /h/ changes”. *Language and Speech* 40:3.249–276.

Gahl, Susanne. 2008. “*Time* and *Thyme* are not homophones: The effect of lemma frequency on word duration in spontaneous speech”. *Language* 84.474–496.

Gries, Stefan & Dagmar Divjak. Forthcoming. *Frequency Effects in Language: Learning and processing*. The Hague: Mouton de Gruyter.

Guy, Gregory. 2005. “Variationist approaches to phonological change”. *The Handbook of Historical Linguistics* ed. by Brian D. Joseph & Richard D. Janda, 369–400. Oxford: Blackwell Publishing.

Hall, Robert A. 1968. “Initial consonants and syntactic doubling in Western Romance”. *Language* 40:4.551–556.

Hopper, Paul & Elizabeth C. Traugott. 2003. *Grammaticalization*. Cambridge, UK: Cambridge University Press.

Jurafsky, Dan, Alan Bell, Michelle Gregory & William D. Raymond. 2001. “Probabilistic relations between words: Evidence from reduction in lexical production”. *Frequency and the Emergence of Linguistic Structure* ed. by Joan Bybee & Paul Hopper, 229–254. Amsterdam: Benjamins.
Labov, William. 1981. “Resolving the Neogrammarian controversy”. *Language* 57.267–308.
Labov, William. 1994. *Principles of Linguistic Change:* Vol. 1, *Internal factors*. Oxford: Blackwell.
Lipski, John. 1995. “[round] and [labial] in Spanish and the ‘free-form’ syllable”. *Linguistics* 33.282–304.
López Chávez, Juan. 1977. “El Fonema /s/ en el Habla de La Cruz, Sinaloa”. *Nueva Revista de Filología Hispánica* 26.332–340.
Malmberg, Bertil. 1958. “Le Pasaje Castillan f > h Perte d’un Trait Redondant”. *Mélanges Petrovici/Cercetari de Lingüística* 3.337–343.
Mason, Keith William. 1994. *Comerse las Eses: A selective bibliographic survey of /s/ aspiration and deletion in dialects of Spanish*. PhD dissertation, University of Michigan.
Méndez Dosuna, Julián. 1996. “Can weakening processes start in initial position?”. *Natural Phonology: The state of the art* ed. by Bernhard Hurch & Richard Rhodes, 97–106. New York: Mouton de Gruyter.
Menéndez Pidal, Ramón. 1968. *Orígenes del Español*. 6a Edición Madrid: Espasa-Calpe.
Naro, Anthony J. 1972. “On ‘f > h’ in Castilian and Western Romance”. *Zeitschrift für Romanische Philologie* 88.435–437.
Paredes, Florentino & Pedro Sánchez-Prieto Borja. 2008. “A methodological approach to the history of the sociolinguistics of the Spanish language”. *International Journal of the Sociology of Language* 193/194.21–55.
Penny, Ralph J. 1972. “The reemergence of /f/ as a phoneme in Castilian”. *Zeitschrift für Romanische Philologie* 88.463–482.
Penny, Ralph J. 1990. “Labiodental /f/, aspiration and /h/-dropping in Spanish: The evolving phonemic values of the graphs f and h”. *Cultures in Contact in Medieval Spain: Historical and literary essays presented to L.P. Harvey (King’s College London Medieval Studies, III)* ed. by David Hook & Barry Taylor, 157–182. London: King’s College.
Penny, Ralph J. 1991. *A History of the Spanish Language*. Cambridge: Cambridge University Press.
Penny, Ralph J. 2000. *Variation and Change in Spanish*. Cambridge: Cambridge University Press.
Pensado, Carmen. 1993. “Sobre el Contexto del Cambio F > h en Castellano”. *Romance Philology* 47:2.147–176.
Phillips, Betty S. 2006. *Word Frequency and Lexical Diffusion*. New York: Palgrave Macmillan.
Pierrehumbert, Janet. 2001. “Exemplar dynamics: Word frequency, lenition, and contrast”. *Frequency and the Emergence of Linguistic Structure* ed. by Joan Bybee & Paul Hopper, 137–157. Amsterdam: Benjamins.

Raymond, William, Robin Dautricourt & Elizabeth Hume. 2006. “Word-medial /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors”. *Language Variation and Change* 18.55–97.
Raymond, William D. & Esther L. Brown. Forthcoming. “Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish”. *Frequency Effects in Language: Learning and Processing* ed. by Stefan Th. Gries & Dagmar S. Divjak. The Hague: Mouton de Gruyter.
Sanchez, Tara. 2008. “Accountability in morphological borrowing: Analyzing a linguistic subsystem as a sociolinguistic variable”. *Language Variation and Change* 20.225–253.
Schuchardt, Hugo. 1885. *Über die Lautgesetze: Gegen die Junggrammatiker*. Berlin: R. Oppenheim.
Sonderegger, Morgan & Partha Niyogiy. 2010. *Variation and Change in English Noun/Verb Pair Stress: Data, dynamical systems models, and their interaction*. Ms., University of Chicago.
Spaulding, Robert. 1943. *How Spanish Grew.* Berkeley: University of California Press.
Timberlake, Alan. 1978. “Uniform and alternating environments in phonological change”. *Folia Slavica* 2.312–328.
Zipf, George K. 1929. “Relative frequency as a determinant of phonetic change”. *Harvard Studies in Classical Philology* 15.1–95.

Résumé
Dans l'utilisation d'un corpus de texte en espagnol médiéval, nous examinons les facteurs affectant le résultat de l’espagnole standard moderne de l'initiale /f/ en latin FV- mots. Les analyses de régression révèlent que la fréquence d'utilisation d'un mot dans les environnements phonétiques extralexicaux et les modes d’emphase lexicale prédit de façon significative la distribution moderne de *f-* ([f]) et *h-* (Ø) dans le lexique espagnol de FV- mots. La quantification de contexte phonétique extralexicale d'usage n'a pas été incorporée dans les études de la phonologie diachronique. Nous ne trouvons aucun effet de la fréquence des mots, de la phonologie lexicale, de la classe de mot, ou de l'histoire de transmission de mot. Les résultats suggèrent que, plutôt que la fréquence d'utilisation, c’est plus specialement la probabilité d’utilisation d'un mot dans des contextes favorisant la réduction qui favorise le changement phonologique. L'échec à trouver un effet significatif de l'histoire de la transmission souligne l'importance relative des sources de changements internes du langage. Les résultats sont compatibles avec les approches basées sur l’usage; les variations contextuelles créent des pressions différentielles articulatoires entre les mots, ce qui donne des prononciations variables qui, lorsqu'elles sont enregistrées dans la mémoire, promeuvent le changement diachronique.

Zusammenfassung
Mit einem Corpus von mittelalterlichen spanischen Texten untersuchen wir Faktoren, die das Ergebnis des initialen /f/ in Lateinischen FV-Wörtern beinflussen. Analysen der Regression zeigen, dass die Frequenz eines Wortbebrauchs in extralexikalischen, phonetisch reduzierenden Umgebungen, und lexikalische Betonungsmuster, die moderne Verteilung der f-([f]) und *h*-(Ø) im spanischen Lexikon der FV-Wörtern aussagekräftig vorhersagen. Die Quantifizierung des extralexikalischen phonetischen

Gebrauchskontextes wurde bisher nicht in Studien der diachronen Phonologie eingearbeitet. Wir finden keinen Effekt durch Wortfrequenz, lexikalische Phonologie, Wortklasse oder Geschichte der Wortübertragung. Die Ergebnisse legen nahe, dass anstatt der Frequenz der Benutzung speziell die Wahrscheinlichkeit der Verwendung eines Wortes in Kontexten, die die Reduktion begünstigen, die phonologische Veränderungen fördert. Das Ausbleiben eines aussagekräftigen Effektes der Übertragungsgeschichte betont die relative Bedeutung einer internen Herkunft der Veränderung in der Sprache. Die Ergebnisse sind im Einklang mit Usage-Based Betrachtungsweisen; kontextuelle Variation erzeugt differentiellen artikulatorischen Zwang zwischen Wörtern, wodurch sich variable Aussprachen bilden, die diachronische Veränderungen bewirken, wenn sie in Erinnerungen registriert sind.

Authors' Addresses

Esther L. Brown
University of Colorado at Boulder
Department of Spanish & Portuguese
McKenna Languages Building, UCB 278
BOULDER, CO 80309

Eshter.Brown@Colorado.edu

William D. Raymond
University of Colorado at Boulder
Department of Psychology and Neuroscience
Muenzinger Psychology Building, UCB 345
BOULDER, CO 80309

William.Raymond@Colorado.edu